Quantum ciphertext authentication and key recycling with the trap code

Dulek, Yfke; Speelman, Florian

Published in:
13th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2018

DOI:
10.4230/LIPIcs.TQC.2018.1

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Quantum Ciphertext Authentication and Key Recycling with the Trap Code

Yfke Dulek
Qusoft, Centrum voor Wiskunde en Informatica, Amsterdam, the Netherlands
dulek@cwi.nl

Florian Speelman[1]
QMATH, Department of Mathematical Sciences, University of Copenhagen, Denmark
speelman@math.ku.dk

Abstract

We investigate quantum authentication schemes constructed from quantum error-correcting codes. We show that if the code has a property called *purity testing*, then the resulting authentication scheme guarantees the integrity of ciphertexts, not just plaintexts. On top of that, if the code is *strong* purity testing, the authentication scheme also allows the encryption key to be recycled, partially even if the authentication rejects. Such a strong notion of authentication is useful in a setting where multiple ciphertexts can be present simultaneously, such as in interactive or delegated quantum computation. With these settings in mind, we give an explicit code (based on the trap code) that is strong purity testing but, contrary to other known strong-purity-testing codes, allows for natural computation on ciphertexts.

2012 ACM Subject Classification Theory of computation → Cryptographic protocols, Theory of computation → Error-correcting codes, Security and privacy → Information-theoretic techniques, Security and privacy → Symmetric cryptography and hash functions, Theory of computation → Quantum information theory

Keywords and phrases quantum authentication, ciphertext authentication, trap code, purity-testing codes, quantum computing on encrypted data

Digital Object Identifier 10.4230/LIPIcs.TQC.2018.1

Related Version `https://arxiv.org/abs/1804.02237` (full version)

Acknowledgements We thank Gorjan Alagic, Christian Majenz, and Christian Schaffner for valuable discussions and useful input at various stages of this research. Aditionally, we thank Christian Schaffner for his comments on an earlier version of this manuscript.

1 Introduction

A central topic in cryptography is authentication: how can we make sure that a message remains unaltered when we send it over an insecure channel? How do we protect a file from being corrupted when it is stored someplace where adversarial parties can potentially access it? And, especially relevant in the current era of cloud computing, how can we let an untrusted third party compute on such authenticated data?

[1] European Research Council (ERC Grant Agreement no 337603), the Danish Council for Independent Research (Sapare Aude), Qubiz Quantum Innovation Center, and VILLUM FONDEN via the QMATH Centre of Excellence (Grant No. 10059)

Following extensive research on authentication of classical data, starting with the seminal work by Wegman and Carter [17], several schemes have been proposed for authenticating quantum states [5, 1, 6]. One notable such scheme is the trap code [6], an encoding scheme that surrounds the data with dummy qubits that function as *traps*, revealing any unauthorized attempts to alter the plaintext data. A client holding the classical encryption key can guide a third party in performing computations directly on the ciphertext by sending input-independent auxiliary quantum states that help bypass the traps, and updating the classical key during the computation. The result is an authenticated output ciphertext.

The trap code distinguishes itself from other quantum authentication schemes in two ways. First, individually-authenticated input qubits can be entangled during the computation, but still be de-authenticated individually. This contrasts for example the Clifford code [1], where de-authentication needs to happen simultaneously on all qubits that were involved in the computation, including any auxiliary ones. Second, the trap code allows for 'authenticated measurements': if a third party measures a ciphertext, the client can verify the authenticity of the result from the classical measurement outcomes only. It is not known how to perform authenticated measurements on other codes. These two qualities make the trap code uniquely suited for quantum computing on authenticated data. It was originally designed for its use in quantum one-time programs [6], but has found further applications in zero-knowledge proofs for QMA [7], and in quantum homomorphic encryption with verification [2].

The extraordinary structure of the trap code is simultaneously its weakness: an adversary can learn information about the secret key by altering the ciphertext in a specific way, and observing whether or not the result is accepted by the client. Thus, to ensure security after de-authentication, the key needs to be refreshed before another quantum state is authenticated. This need for a refresh inhibits the usefulness of the trap code, because computation on multiple qubits under the trap code requires these qubits to be authenticated under overlapping secret keys.

In recent years, several works have refined the original definition of quantum authentication by Barnum et al. [5]. The trap code is secure under the weakest of these definitions [10], where only authenticity of the plaintext is guaranteed. But, as argued, it is not under the stronger 'total authentication' [12], where no information about the key is leaked if the client accepts the authentication. As Portmann mentions in his work on authentication with key recycling in the abstract-cryptography framework [15], it is not even clear whether the trap code can be regarded as a scheme with *partial* key leakage, as defined in [12], because of the adaptive way in which it can be attacked. In a different direction, Alagic, Gagliardoni, and Majenz [3] define a notion of quantum ciphertext authentication (QCA), where also the integrity of the ciphertext is guaranteed, and not just that of the plaintext. Ciphertext authentication is incomparable with total authentication: neither one implies the other. Before the current work, it was unknown whether the trap code authenticates ciphertexts.

Barnum et al. [5] built schemes for authentication of quantum data based on quantum error-correcting codes that are *purity testing*, meaning that any bit or phase flip on the message is detected with high probability. Portmann [15], working in the abstract-cryptography framework, showed that if the underlying code satisfies a stronger requirement called *strong purity testing*, the resulting authentication scheme allows for complete key recycling in the accept case, and for partial key recycling in the reject case. The trap code can be seen as a purity-testing error-correcting code, but it is not strong purity testing. This is consistent with the observation that keys in the trap code cannot be recycled.

Quantum plaintext authentication with key recycling has been studied before. Oppenheim and Horodecki [14] showed partial key recycling for schemes based on purity testing codes,

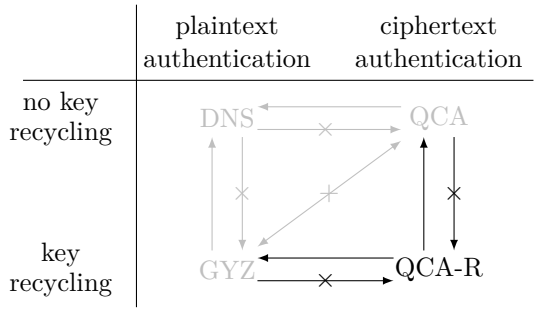

Figure 1 Overview of different definitions of quantum authentication. Three previously defined notions (in gray) and their relations were already known: DNS [10] is strictly weaker than GYZ [12] (total authentication) and QCA [3]. These last two are incomparable: there exist schemes that satisfy either one, but not the other. On the bottom right, our new definition QCA-R is displayed: it is strictly stronger than both GYZ and QCA. For justifications of the relations displayed in this figure, refer to pages 6 (for DNS $\nrightarrow$ GYZ), 6 (for DNS $\nrightarrow$ QCA), 6 (for GYZ $\nleftrightarrow$ QCA), and 9 (for GYZ $\nrightarrow$ QCA-R and QCA $\nrightarrow$ QCA-R).

under a weaker notion of security. Hayden, Leung, and Mayers [13] adapted Barnum et al.'s construction to use less key and show its authenticating properties in the universal-composability framework. Fehr and Salvail [11] develop a quantum authentication scheme for classical messages that achieves the same key-recycling rate as Portmann [15], but is not based on quantum error-correction and only requires the client to prepare and measure.

1.1 Our contributions

We investigate the relation between (strong) purity testing and quantum ciphertext authentication (QCA), and give a variation on the trap code with stronger security guarantees. We specify our contributions in more detail below.

Section 3: Definition of quantum ciphertext authentication with key recycling *(QCA-R)*. We give a new definition for quantum authentication, QCA-R, that provides both ciphertext authentication and key recycling, and is thereby strictly stronger than existing definitions. See Figure 1 for a comparison of different notions of authentication.

Section 3.1: Purity-testing codes give rise to QCA-secure encryption. We prove that Barnum et al.'s canonical construction of authentication schemes from purity-testing codes [5] produces schemes that are not only plaintext authenticating, but also ciphertext authenticating (QCA). The proof generalizes the proofs in [8] that the trap code and Clifford code are plaintext authenticating, using a different (but still efficient) simulator. Note that our result immediately implies that the trap code is ciphertext authenticating.

Section 3.1: Strong-purity-testing codes give rise to QCA-R-secure encryption. Purity-testing codes are generally not sufficient for constructing QCA-R schemes, but strong-purity-testing codes are: we prove that Barnum et al.'s canonical construction achieves QCA-R when a strong-purity-testing code is used as a resource. In case the authenticated message is accepted, the entire key can be reused. Otherwise, all but the quantum-one-time-pad key can be reused.

Section 4: A strong-purity-testing version of the trap code. We give an explicit construction of a strong-purity-testing code that is inspired by the trap code. In this *strong trap code*, the underlying error-correcting code is not only applied to the data qubits, but also to the trap qubits. The result is a quantum authentication scheme which satisfies the strong notion of QCA-R, but still maintains the computational properties that make the original trap code such a useful scheme.

Section 5: Security under parallel encryption. To illustrate the power of recycling key in the reject case, we consider a setting with a different type of key reuse: reusing (part of) a key immediately to authenticate a second qubit, even before the first qubit is verified. We show that, if multiple qubits are simultaneously authenticated using a scheme that is based on a strong-purity-testing code, then de-authenticating some of these qubits does not jeopardize the security of the others, even if their keys overlap. This property is especially important when using the computational capabilities of the strong trap code, since computing on authenticated qubits needs multiple qubits to use overlapping keys.

2 Preliminaries

2.1 Notation

We use conventional notation for unitary matrices (U or V), pure states ($|\psi\rangle$ or $|\varphi\rangle$), and mixed states (ρ or σ). We reserve the symbol τ for the completely mixed state $\mathbb{I}/d$, and $|\Phi^+\rangle$ for the EPR pair $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. The m-qubit Pauli group is denoted with $\mathbb{P}_m$, and its elements with P_ℓ where ℓ is a $2m$-bit string indicating the bit-flip and phase-flip positions. By convention, P_0 is identity. The X-weight (, Y-weight, or Z-weight) of a Pauli is the number of qubits on which it acts as an X (, Y, or Z) operation.

We often specify the register(s) on which a unitary acts by gray superscripts (as in U^R); it is implicit that the unitary acts as identity on all other registers. The trace norm of a density matrix ρ is written as $\|\rho\|_{\mathrm{tr}}$. The diamond norm of a channel Ψ is written as $\|\Psi\|_\diamond := \sup_\rho \|(\mathbb{I}\otimes\Psi)(\rho)\|_{\mathrm{tr}}$. If we want to talk about the distance between two channels Ψ and Ψ', we use the normalized quantity $\frac{1}{2}\|\Psi-\Psi'\|_\diamond$, which we refer to as the *diamond-norm distance*.

2.2 Quantum authentication

A secret-key quantum encryption scheme consists of three (efficient) algorithms: key generation KeyGen, encryption $\mathsf{Encrypt}_k$, and decryption $\mathsf{Decrypt}_k$. Throughout this work, we will assume that KeyGen selects a key k uniformly at random from some set $\mathcal{K}$; our results still hold if the key is selected according to some other distribution. By Lemma B.9 in [4], we can characterize the encryption and decryption maps as being of the form

$$\mathsf{Encrypt}_k : \rho^M \mapsto U_k^{MT}(\rho\otimes\sigma_k^T)(U_k^\dagger)^{MT}, \tag{1}$$

$$\mathsf{Decrypt}_k : \rho^{MT} \mapsto \mathrm{Tr}_T\left[(\Pi_k^{\mathsf{acc}})^T\left(U_k^\dagger\rho U_k^{MT}\right)(\Pi_k^{\mathsf{acc}})^T\right] + D_k^{MT}\left[(\Pi_k^{\mathsf{rej}})^T\left(U_k^\dagger\rho U_k^{MT}\right)(\Pi_k^{\mathsf{rej}})^T\right]. \tag{2}$$

Here, M is the message register, σ_k is some key-dependent *tag* state in register T, and U_k is a unitary acting on both. Π_k^{acc} and Π_k^{rej} are orthogonal projectors onto the support of σ_k and its complement, respectively. Finally, D_k is any channel: we will usually assume that $D_k(\cdot) = \mathrm{Tr}_{MT}(\cdot)\otimes|\perp\rangle\langle\perp|^M$, i.e., it traces out the message and tag register entirely,

and replaces the message with some dummy state that signifies a reject. Because of the above characterization, we will often talk about encryption schemes as a keyed collection $\{(U_k, \sigma_k)\}_{k\in\mathcal{K}}$ of unitaries and tag states.

There are several definitions of the authentication of quantum data. All definitions involve some parameter ε; unless otherwise specified, we require ε to be negligibly small in the size of the ciphertext.

The simplest definition is that of plaintext authentication, presented in [10] (although their definition was in phrased terms of the trace norm), where no guarantees are given about the recyclability of the key.

▶ **Definition 1** (Quantum plaintext authentication (DNS) [10])**.** A quantum encryption scheme $\{(U_k, \sigma_k)\}_{k\in\mathcal{K}}$ is *plaintext ε-authenticating* (or ε-DNS) if for all CP maps $\mathcal{A}$ (acting on the message register M, tag register T, and a side-information register R), there exist CP maps $\mathcal{S}_{\mathsf{acc}}$ and $\mathcal{S}_{\mathsf{rej}}$ such that $\mathcal{S} := \mathcal{S}_{\mathsf{acc}} + \mathcal{S}_{\mathsf{rej}}$ is trace-preserving, and

$$\frac{1}{2}\left\| \mathop{\mathbb{E}}_{k} \left[\mathsf{Decrypt}_k \circ \mathcal{A}^{MTR} \circ \mathsf{Encrypt}_k\right]^{MR} - \left(\mathbb{I}^M \otimes \mathcal{S}_{\mathsf{acc}}^R + |\bot\rangle\langle\bot|^M (\mathrm{Tr}_M \otimes \ \mathcal{S}_{\mathsf{rej}}^R)\right)\right\|_\diamond \leq \varepsilon,$$

where $\mathsf{Encrypt}_k$ and $\mathsf{Decrypt}_k$ are of the form of equations (1) and (2).

The simulator in Definition 1 reflects the ideal functionality of an authentication scheme: in the accept case, the message remains untouched, whereas in the reject case, it is completely discarded and replaced with the fixed state $|\bot\rangle\langle\bot|$. Any action on the side-information register R is allowed.

▶ **The trap code.** An example of a plaintext-authenticating scheme is the trap code [6]. This scheme encrypts single-qubit messages by applying a fixed distance-d CSS code E to the message, producing n physical qubits, and then appending $2n$ "trap" qubits (n computational-basis traps in the state $|0\rangle\langle 0|$, and n Hadamard-basis traps in the state $|+\rangle\langle +|$). The resulting $3n$ qubits are permuted in a random fashion according to a key k_1, and one-time padded with a second key k_2. At decryption, the one-time pad and permutation are removed, the traps are measured in their respective bases, and the syndrome of the CSS code is checked.[2] The trap code, for a key $k = (k_1, k_2)$, is characterized by $U_k = P_{k_2}\pi_{k_1}(E \otimes \mathsf{I}^{\otimes n} \otimes \mathsf{H}^{\otimes n})$ and $\sigma_k = |0\rangle\langle 0|^{\otimes(3n-1)}$, where π_{k_1} is a unitary that permutes the $3n$ qubits. A proof that the trap code is plaintext $(2/3)^{d/2}$-authenticating can be found in e.g. [8].

Another definition of quantum authentication is presented in [12] (where it is called 'total authentication'): in this definition, the key should be recyclable in the accept case. This is modeled by revealing the key to the environment after use, and requiring that it is indistinguishable from a completely fresh and uncorrelated key. If that is the case, it can be recycled for another round.

▶ **Definition 2** (Quantum plaintext authentication with key recycling (GYZ) [12])**.** A quantum encryption scheme $\{(U_k, \sigma_k)\}_{k\in\mathcal{K}}$ is *plaintext ε-authenticating with key recycling* (or ε-GYZ) if for all CP maps $\mathcal{A}$ (acting on the message register M, tag register T, and a side-information register R), there exist CP maps $\mathcal{S}_{\mathsf{acc}}$ and $\mathcal{S}_{\mathsf{rej}}$ such that $\mathcal{S} := \mathcal{S}_{\mathsf{acc}} + \mathcal{S}_{\mathsf{rej}}$ is trace preserving,

[2] We differ from the analysis by Broadbent and Wainewright [8] in that we consider the variant that uses error detection instead of error correction on the data qubits.

and

$$\frac{1}{2}\left\| \mathop{\mathbb{E}}_{k} \left[\rho^{MR} \mapsto \mathrm{Tr}_T \left(\Pi_k^{\mathsf{acc}} U_k^\dagger \left(\mathcal{A}^{MTR}\left(U_k(\rho \otimes \sigma_k^T)U_k^\dagger\right)\right) U_k \Pi_k^{\mathsf{acc}}\right) \otimes |k\rangle\langle k|\right] - \left(\mathbb{I}^M \otimes \mathcal{S}_{\mathsf{acc}}^R \otimes \tau_{\mathcal{K}}\right) \right\|_\diamond \leq \varepsilon.$$

Note that Definition 2 only specifies what should happen in the accept case. Nevertheless, it is a strictly stronger definition than DNS authentication [4].

The trap code is not plaintext ε-authenticating with key recycling for sub-constant ε. To see this, consider an adversary $\mathcal{A}$ that applies X to (only) the first qubit of the MT register. With probability $2/3$, the attack lands on a data qubit or a $|0\rangle\langle 0|$ trap, and is detected. Thus, in the real accept scenario, the key register will contain a mixture of only those keys that permute a $|+\rangle\langle +|$ into the first position. All other keys are diminished by the projector Π_k^{acc}. Since the ideal scenario contains a mixture of *all* possible keys in the key register, the difference between the two channels is considerable. In practice, if an adversary learns whether the authentication succeeded, she gets information about the positions of the traps.

▶ **The Clifford code.** A simple yet powerful code that authenticates plaintexts with key recycling is the Clifford code [1]. In this code, we fix a parameter t, and set $\sigma_k = |0^t\rangle\langle 0^t|$ for all k, and U_k a uniformly random Clifford on $t+1$ qubits. The Clifford code (and any authentication code that is based on a 2-design) is plaintext ε-authenticating with key recycling for $\varepsilon = O(2^{-t})$ [4].

Strengthening Definition 1 in a different direction, Alagic, Gagliardoni, and Majenz recently introduced the notion of quantum ciphertext authentication [3]. This notion does not limit the amount of key leaked, but requires that if authentication accepts, the entire *ciphertext* was completely untouched. Ciphertext authentication is used as an ingredient for quantum encryption that is secure against chosen-ciphertext attacks [3].

▶ **Definition 3** (Quantum ciphertext authentication (QCA) [3])**.** A quantum encryption scheme $\{(U_k, \sigma_k = \sum_r p_{k,r}|\varphi_{k,r}\rangle\langle\varphi_{k,r}|)\}_{k\in\mathcal{K}}$ is *ciphertext ε-authenticating* (or ε-QCA) if it is plaintext ε-authenticating as in Definition 1, and the accepting simulator $\mathcal{S}_{\mathsf{acc}}$ is of the form

$$\mathcal{S}_{\mathsf{acc}} : \rho^R \mapsto \mathop{\mathbb{E}}_{k',r} \left[\langle\varphi_{k',r}|^T \langle\Phi^+|^{M_1M_2} U_{k'}^\dagger \mathcal{A}^{M_1TR}\left(U_{k'}^{M_1T} \rho_{k',r}^{RM_1M_2T} U_{k'}^\dagger\right) U_{k'}|\varphi_{k',r}\rangle|\Phi^+\rangle\right].$$

where $\rho_{k',r} := \rho^R \otimes |\Phi^+\rangle\langle\Phi^+|^{M_1M_2} \otimes |\varphi_{k',r}\rangle\langle\varphi_{k',r}|^T$ is the input state before (simulated) encryption.

In QCA, the accepting simulator tests whether the message remains completely untouched by encrypting half of an EPR pair (stored in register M_1) as a 'dummy message', under a key k' that it generates itself. It remembers the randomness r used in creating the tag state σ_k, so that it can test very accurately whether the tag state was untouched. Because $\mathcal{S}_{\mathsf{acc}}$ remembers the randomness, a scheme that appends a qubit at the end of its ciphertexts, but never checks its state at decryption time, cannot be ciphertext authenticating. The Clifford code *is* QCA [3], as is the trap code (see Section 3.1).

In general, key recycling as in Definition 2 does not imply QCA. To see this, take any scheme $\{(U_k, \sigma_k)\}_{k\in\mathcal{K}}$ that is plaintext authenticating with key recycling, and alter it by appending a qubit in the fully mixed state to σ_k (and extending U_k to act as identity on this qubit). This scheme still satisfies Definition 2, but cannot be ciphertext authenticating,

because attacks on this last qubit are not noticed in the real scenario. Conversely, not all ciphertext-authenticating schemes have key recycling. Take any scheme that is QCA, and alter it by adding one extra bit b of key, and setting $\sigma_{kb} := \sigma_k \otimes |b\rangle\langle b|$ and $U_{kb} := U_k \otimes \mathsf{I}$, effectively appending the bit of key at the end of the ciphertext. This scheme still satisfies Definition 3, but leaks at least one bit of key.[3] For an overview of the relations between DNS, GYZ, and QCA, refer to Figure 1 on page 3.

2.3 (Strong) purity testing in quantum error correction

An $[[n, m]]$ quantum error-correcting code (QECC), characterized by a unitary operator V, encodes a message ρ consisting of m qubits into a codeword $V(\rho \otimes |0^t\rangle\langle 0^t|)V^\dagger$ consisting of n qubits, by appending $t := n - m$ tags $|0\rangle\langle 0|$, and applying the unitary V. Decoding happens by undoing the unitary V, and measuring the tag register in the computational basis. The measurement outcome is called the syndrome: an all-zero syndrome indicates that no error-correction is necessary. In this work, we will only use the error-detection property of QECCs, and will not worry about how to correct the message if a non-zero syndrome is measured. If that happens, we will simply discard the message (i.e., reject).

For any bit string $x \in \{0,1\}^m$, let $|x_L\rangle$ (for "logical $|x\rangle$") denote a valid encoding of $|x\rangle$, i.e., a state that will decode to $|x\rangle$ without error. A defining feature of any QECC is its distance: the amount of bit and/or phase flips required to turn one valid codeword into another. If we want to be explicit about the distance d of an $[[n, m]]$ code, we will refer to it as an $[[n, m, d]]$ code.

▶ **Definition 4** (Distance). The *distance* of an $[[n, m]]$ code is the minimum weight of a Pauli P such that $P|x_L\rangle = |y_L\rangle$ for some $x \neq y$, with $x, y \in \{0,1\}^m$.

In a cryptographic setting, it can be useful to select a code from a set of codes $\{V_k\}_{k\in\mathcal{K}}$ for some key set $\mathcal{K}$. We will again assume that the key k is selected uniformly at random.

Following [5] and [15], we restrict our attention to codes for which applying a Pauli to a codeword is equivalent to applying a (possibly different) Pauli directly to the message and tag register. In other words, the unitary V must be such that for any $P_\ell \in \mathbb{P}_{m+t}$, there exists a $P_{\ell'} \in \mathbb{P}_{m+t}$ and a $\theta \in \mathbb{R}$ such that $P_\ell V = e^{i\theta} V P_{\ell'}$. With our attention restricted to codes with this property, we can meaningfully define the following property:

▶ **Definition 5** (Purity testing [5]). A set of codes $\{V_k\}_{k\in\mathcal{K}}$ is *purity testing* with error ε if for any Pauli $P_\ell \in \mathbb{P}_{m+t} \backslash \{\mathsf{I}^{\otimes(m+t)}\}$,

$$\Pr_k \left[V_k^\dagger P_\ell V_k \in (\mathbb{P}_m \backslash \{\mathsf{I}^{\otimes m}\}) \otimes \{\mathsf{I}, \mathsf{Z}\}^{\otimes t} \right] \leq \varepsilon.$$

In words, for any non-identity Pauli, the probability (over the key) that the Pauli alters the message but is not detected (i.e., no tag bit is flipped) is upper bounded by ε.

The trap code (see page 5) based on an $[[n, 1, d]]$ CSS code, without the final quantum one-time pad, is a purity-testing code with error $(2/3)^{d/2}$ [6]. In our framework, the trap code is described as a QECC with $m = 1$, $t = 3n - 1$, and $V_k = \pi_k(E \otimes \mathsf{I}^{\otimes n} \otimes \mathsf{H}^{\otimes n})$.

Note that purity-testing codes do not necessarily detect *all* Pauli attacks with high probability: it may well be that a Pauli attack remains undetected, because it acts as identity on the message. Flipping the first bit of a trap-code ciphertext is an example of

[3] We thank Gorjan Alagic and Christian Majenz for providing these example schemes that show the separation between Definitions 2 and 3.

such an attack: it remains undetected with probability $1/3$ (if it hits a $|+\rangle$ trap), but unless it is detected, it also does not alter the message. An attacker may use this fact to learn information about the permutation π_k by observing whether or not the QECC detects an error.

The above exploitation of purity-testing codes has led Portmann to consider a stronger notion of purity testing that should allow for keys to be safely reusable. In this definition, even the Paulis that act as identity on the message should be detected:

▶ **Definition 6** (Strong purity testing [15])**.** A set of codes $\{V_k\}_{k\in\mathcal{K}}$ is *strong purity testing* with error ε if for any Pauli $P_\ell \in \mathbb{P}_{m+t}\backslash\{\mathsf{I}^{\otimes(m+t)}\}$,

$$\Pr_k\left[V_k^\dagger P_\ell V_k \in \mathbb{P}_m \otimes \{\mathsf{I},\mathsf{Z}\}^{\otimes t}\right] \leq \varepsilon.$$

The Clifford code is strong purity testing with error 2^{-t}, as is any other unitary 2-design [15]. As informally discussed above, the trap code is not strong purity testing for any small ε.

Barnum et al. [5] described a canonical method of turning a QECC set $\{V_{k_1}\}_{k_1\in\mathcal{K}_1}$ into a symmetric-key encryption scheme. The encryption key k consists of two parts: the key $k_1 \in \mathcal{K}_1$ for the QECC, and an additional one-time pad key $k_2 \in \{0,1\}^{2(m+t)}$. The encryption map is then defined by setting $U_{k_1,k_2} := P_{k_2}V_{k_1}$, and $\sigma_{k_1,k_2} = |0^t\rangle\langle 0^t|$. Since σ_{k_1,k_2} is key-independent, the projectors $\Pi^{\mathsf{acc}} = |0^t\rangle\langle 0^t|$ and $\Pi^{\mathsf{rej}} = \mathbb{I} - |0^t\rangle\langle 0^t|$ are key-independent as well. In Construction 1, the complete protocol is described. In [6], protocols of this form are called "encode-encrypt schemes".

Construction 1 Barnum et al.'s canonical construction [5] of a symmetric-key encryption scheme from an $[[m+t,m]]$ quantum error-correcting code $\{V_{k_1}\}_{k_1\in\mathcal{K}_1}$.

`Generate keys:` sample $k_1 \leftarrow \mathcal{K}_1$ and $k_2 \leftarrow \mathcal{K}_2 = \{0,1\}^{2(m+t)}$.
`Encrypt:` $\rho^M \mapsto P_{k_2}^{MT} V_{k_1}^{MT}(\rho^M \otimes |0^t\rangle\langle 0^t|^T) V_{k_1}^{MT\dagger} P_{k_2}^{MT\dagger}$.
`Decrypt:` $\rho^{MT} \mapsto \langle 0^t|\left(V_{k_1}^\dagger P_{k_2}^\dagger \rho P_{k_2} V_{k_1}\right)|0^t\rangle + |\perp\rangle\langle\perp|^M \otimes \mathrm{Tr}_M\left[\sum_{i\neq 0^t}\langle i|\left(V_{k_1}^\dagger P_{k_2}^\dagger \rho P_{k_2} V_{k_1}\right)|i\rangle\right]$

When using Construction 1 with a strong-purity-testing code, plaintext authentication with key recycling is achieved, even with partial key recycling in the reject case [15]. If just a purity-testing code is used for the construction, the resulting encryption scheme is plaintext authenticating [5], but not necessarily with key recycling (the trap code is a counterexample).

3 Quantum ciphertext authentication with key recycling (QCA-R)

In this section, we will define a notion of quantum authentication that is stronger than all of Definitions 1, 2, and 3. We will show that Construction 1, when used with a strong-purity-testing code, results in an authentication scheme in this new, stronger sense.

▶ **Definition 7** (Quantum ciphertext authentication with key recycling (QCA-R))**.** A quantum encryption scheme $\{(U_k, \sigma_k = \sum_r p_{k,r}|\varphi_{k,r}\rangle\langle\varphi_{k,r}|)\}_{k\in\mathcal{K}}$ is *ciphertext ε-authenticating with key recycling* (or ε-QCA-R), with key recycling function f, if for all CP maps $\mathcal{A}$ (acting on the message register M, tag register T, and a side-information register R), there exists a CP map $\mathcal{S}_{\mathsf{rej}}$ such that

$$\begin{aligned}\mathfrak{R} : \rho^{MR} \mapsto \mathbb{E}_k\Big[&\mathrm{Tr}_T\left(\Pi^{acc}\left(U_k^\dagger \mathcal{A}^{MTR}\left(U_k^{MT}(\rho\otimes\sigma_k^T)U_k^{\dagger}\right)U_k\right)\Pi^{acc}\right)\otimes|k\rangle\langle k| \\ &+ |\perp\rangle\langle\perp|^M \otimes \mathrm{Tr}_{MT}\left(\Pi^{rej}\left(U_k^\dagger \mathcal{A}^{MTR}\left(U_k^{MT}(\rho\otimes\sigma_k^T)U_k^{\dagger}\right)U_k\right)\Pi^{rej}\right)\otimes|f(k)\rangle\langle f(k)|\Big]\end{aligned}$$

is ε-close in diamond-norm distance to the ideal channel,

$$\mathfrak{I} : \rho^{MR} \mapsto \left(\mathbb{I}^M \otimes \mathcal{S}^{\mathsf{acc}}\right)(\rho^{MR}) \otimes \tau_{\mathcal{K}} \quad + \quad |\perp\rangle\langle\perp|^M \otimes \ \mathcal{S}^{\mathsf{rej}}(\rho^R) \otimes \mathbb{E}_k \left[|f(k)\rangle\langle f(k)|\right],$$

where $\mathcal{S} := \mathcal{S}_{\mathsf{acc}} + \mathcal{S}_{\mathsf{rej}}$ is trace preserving, and $\mathcal{S}_{\mathsf{acc}}$ is as in Definition 3 of QCA, that is,

$$\mathcal{S}_{\mathsf{acc}} : \rho^R \mapsto \mathop{\mathbb{E}}_{k',r} \left[\langle\varphi_{k',r}|^T \langle\Phi^+|^{M_1M_2} U_{k'}^\dagger \mathcal{A}^{M_1TR}\left(U_{k'}^{M_1T} \rho_{k',r}^{RM_1M_2T} U_{k'}^\dagger\right) U_{k'}|\varphi_{k',r}\rangle|\Phi^+\rangle\right]$$

for $\rho_{k',r} := \rho^R \otimes |\Phi^+\rangle\langle\Phi^+|^{M_1M_2} \otimes |\varphi_{k',r}\rangle\langle\varphi_{k',r}|^T$.

The first condition (closeness of the real and ideal channel) is a strengthening of Definition 2: following Portmann [15], we also consider which part of the key can be recycled in the reject case. If the recycling function f is the identity function, all of the key can be recycled. If f maps all keys to the empty string, then no constraints are put on key leakage in the reject case.

QCA-R strengthens both GYZ and QCA, but not vice versa: the schemes from Section 2.2 that separate the two older notions are immediately examples of schemes that are GYZ or QCA but cannot be QCA-R.

3.1 Constructing QCA-R from any strong-purity-testing code

It was already observed that if a set of quantum error-correcting codes $\{V_{k_1}\}_{k_1 \in \mathcal{K}_1}$ is purity testing, then the encryption scheme resulting from Construction 1 is plaintext authenticating [5]. We strengthen this result by showing that the construction turns purity-testing codes into *ciphertext*-authenticating schemes (Theorem 8), and strong-purity-testing codes into QCA-R schemes (Theorem 9). Only purity testing is in general not enough to achieve QCA-R: the trap code is again a counterexample.

▶ **Theorem 8.** *Let $\{V_{k_1}\}_{k_1 \in \mathcal{K}_1}$ be a purity-testing code with error ε. The encryption scheme resulting from Construction 1 is quantum ciphertext ε-authenticating (ε-QCA).*

Sketch. In order to prove Theorem 8, we define a simulator that runs the adversary on encrypted halves of EPR pairs, so that the simulator is of the correct form for Definition 3. We prove that the ideal and the real channel are close by considering the accept and the reject cases separately, and by showing that they are both $\varepsilon/2$-close. First, we decompose the adversarial attack into Paulis by Pauli twirling [9] it with the quantum-one-time-pad encryption from Construction 1. In the accept case, the difference between the real and the ideal scenario lies in those attacks that are accepted in the real case, but not in the ideal case. These are exactly those Paulis that, after conjugation with the key k_1 that indexes the purity-testing code, are in the set $(\mathbb{P}_m \otimes \{\mathsf{I},\mathsf{Z}\}^{\otimes t}) \backslash (\{\mathsf{I}^{\otimes m}\} \otimes \{\mathsf{I},\mathsf{Z}\}^{\otimes t}) = (\mathbb{P}_m \backslash \{\mathsf{I}^{\otimes m}\}) \otimes \{\mathsf{I},\mathsf{Z}\}^{\otimes t}$. The purity-testing property guarantees that the probability over k_1 of a Pauli attack landing in this set is small. The reject case is similar. ◀

A full proof of Theorem 8 can be found in the full version. The proof of Theorem 9 below uses the same techniques. It follows the proof structure of [15, Theorem 3.5], but with a simulator that is suitable for QCA-R.

▶ **Theorem 9.** *Let $\{V_{k_1}\}_{k_1 \in \mathcal{K}_1}$ be a strong-purity-testing code with error ε. The encryption scheme resulting from Construction 1 is quantum ciphertext $(\sqrt{\varepsilon} + \frac{3}{2}\varepsilon)$-authenticating with key recycling (ε-QCA-R), with recycling function $f(k_1, k_2) := k_1$.*

Proof. Let $\mathcal{A}$ be an adversary as in Definition 7. Define a simulator $\mathcal{S}$ on the side-information register R as follows: prepare an EPR pair $|\Phi^+\rangle\langle\Phi^+|$ in the register M_1M_2 and encrypt the first qubit using a freshly sampled key $(k_1', k_2') \in \mathcal{K} := \mathcal{K}_1 \times \mathcal{K}_2$ (that is, initialize the tag register T in the state $|0^t\rangle\langle 0^t|$, and apply $P_{k_2'}V_{k_1'}$ to M_1T). Then, run the adversary on the registers M_1TR, keeping M_2 to the side. Afterwards, run the decryption procedure by undoing the encryption unitary and measuring whether the registers M_1M_2T are still in the state $|\Phi^+, 0^t\rangle\langle\Phi^+, 0^t|$ $(= |\Phi^+\rangle\langle\Phi^+| \otimes |0^t\rangle\langle 0^t|)$. If so, accept, and if not, reject. Note that this simulator is of the required form in the accept case (see Definition 7).

We show that for this simulator, the distance $\frac{1}{2}\|\mathfrak{I} - \mathfrak{R}\|_\diamond$ between the ideal and the real channel is upper bounded by $\sqrt{\varepsilon} + \frac{3}{2}\varepsilon$. Let ρ^{MRE} be any quantum state on the message register, side-information register, and an environment register E. Let U^{MTR} be a unitary[4] map representing the adversarial channel $\mathcal{A}$, and let $\mu^{\mathsf{real}}_{k_1,k_2}$ and $\mu^{\mathsf{ideal}}_{k_1,k_2}$ be the effective output states in the real and ideal world, respectively:

$$\mu^{\mathsf{real}}_{k_1,k_2} := V_{k_1}^\dagger P_{k_2}^\dagger U^{MTR} P_{k_2}^{MT} V_{k_1}^{MT} (\rho \otimes |0^t\rangle\langle 0^t|) V_{k_1}^\dagger P_{k_2}^\dagger U^\dagger P_{k_2} V_{k_1}, \tag{3}$$

$$\mu^{\mathsf{ideal}}_{k_1,k_2} := V_{k_1}^\dagger P_{k_2}^\dagger U^{M_1TR} P_{k_2}^{M_1T} V_{k_1}^{M_1T} (\rho \otimes |0^t, \Phi^+\rangle\langle 0^t, \Phi^+|) V_{k_1}^\dagger P_{k_2}^\dagger U^\dagger P_{k_2} V_{k_1}. \tag{4}$$

Then we can write the result of the real and the ideal channels as

$$\begin{aligned}\mathfrak{R}(\rho) = \mathop{\mathbb{E}}_{k_1,k_2} \Bigg[& \langle 0^t|^T \mu^{\mathsf{real}}_{k_1,k_2} |0^t\rangle \otimes |k_1k_2\rangle\langle k_1k_2| \\ & + |\bot\rangle\langle\bot|^M \otimes \operatorname{Tr}_M \left(\sum_{i \neq 0^t} \langle i|^T \mu^{\mathsf{real}}_{k_1,k_2} |i\rangle \right) \otimes |k_1\rangle\langle k_1| \Bigg],\end{aligned} \tag{5}$$

$$\begin{aligned}\mathfrak{I}(\rho) = \mathop{\mathbb{E}}_{k_1',k_2'} \Bigg[& \langle \Phi^+, 0^t|^{M_1M_2T} \mu^{\mathsf{ideal}}_{k_1',k_2'} |\Phi^+ 0^t\rangle \otimes \tau_{\mathcal{K}} \\ & + |\bot\rangle\langle\bot|^M \otimes \operatorname{Tr}_M \left(\sum_{i \neq (\Phi^+, 0^t)} \langle i|^{M_1M_2T} \mu^{\mathsf{ideal}}_{k_1',k_2'} |i\rangle \right) \otimes \tau_{\mathcal{K}_1} \Bigg].\end{aligned} \tag{6}$$

These expressions are obtained simply by plugging in the description of the authentication scheme (see Construction 1) and the simulator into the channels of Definition 7. Since the accept states are orthogonal to the reject states in the M register, and since the key states are all mutually orthogonal, the distance $\frac{1}{2}\|\mathfrak{I}(\rho) - \mathfrak{R}(\rho)\|_{\mathrm{tr}}$ can be written as

$$\begin{aligned}\mathop{\mathbb{E}}_{k_1,k_2} \frac{1}{2} \left\| \mathop{\mathbb{E}}_{k_1',k_2'} \left(\langle \Phi^+, 0^t | \mu^{\mathsf{ideal}}_{k_1',k_2'} | \Phi^+, 0^t \rangle \right) - \langle 0^t | \mu^{\mathsf{real}}_{k_1,k_2} | 0^t \rangle \right\|_{\mathrm{tr}} \\ + \mathop{\mathbb{E}}_{k_1} \frac{1}{2} \left\| \mathop{\mathbb{E}}_{k_1',k_2'} \left(\operatorname{Tr}_M \sum_{i \neq (0^t, \Phi^+)} \langle i | \mu^{\mathsf{ideal}}_{k_1',k_2'} | i \rangle \right) - \mathop{\mathbb{E}}_{k_2} \left(\operatorname{Tr}_M \sum_{i \neq 0^t} \langle i | \mu^{\mathsf{real}}_{k_1,k_2} | i \rangle \right) \right\|_{\mathrm{tr}}.\end{aligned} \tag{7}$$

For a complete derivation, see the full version. We can thus focus on bounding the two terms in equation (7), for accept and reject, separately. Note the difference between the two terms: in the reject case, the expectation over the one-time pad key k_2 does not have to

[4] We can assume unitarity without loss of generality: if the adversary's actions are not unitary, we can dilate the channel into a unitary one by adding another environment and tracing it out afterwards. In the proof, the environment takes on the same role as the side-information register R, so we omit it for simplicity.

be brought outside of the trace norm, since it is not recycled after a reject. This will make bounding the second term in equation (7) the simpler of the two, so we will start with that one.

Decompose the attack as $U^{MTR} = \sum_\ell \alpha_\ell P_\ell^{MT} \otimes U_\ell^R$. Intuitively, the two states inside the second trace norm differ on those Paulis P_ℓ that are rejected in the ideal scenario, but not in the real one. The strong-purity-testing property promises that these Paulis are very few. However, we have to be careful, because the simulator independently generates its own set of keys. We will now bound the second term in equation (7) more formally.

By rearranging sums, commuting Paulis, and applying projectors (for details: see the full version), we can rewrite the second term inside the trace norm, the state in the real reject case for k_1, as

$$\mathop{\mathbb{E}}_{k_2}\left(\mathrm{Tr}_M \sum_{i\neq 0^t} \langle i|\mu^{\mathsf{real}}_{k_1,k_2}|i\rangle\right) = \mathrm{Tr}_M\left(\sum_{\ell\,:\,V_{k_1}^\dagger P_\ell V_{k_1}\notin\mathbb{P}_{\mathsf{real}}} |\alpha_\ell|^2\, U_\ell^R \rho^{MR} U_\ell^\dagger\right), \tag{8}$$

where $\mathbb{P}_{\mathsf{real}}$ contains the Paulis that are accepted by the real projector, i.e., $\mathbb{P}_{\mathsf{real}} := \mathbb{P}_m \otimes \{\mathsf{I},\mathsf{Z}\}^{\otimes t}$. Similarly, defining $\mathbb{P}_{\mathsf{ideal}} := \{\mathsf{I}^{\otimes m}\} \otimes \{\mathsf{I},\mathsf{Z}\}^{\otimes t}$ to be the set of Paulis that are allowed by the ideal projector, the resulting state in the reject case is

$$\mathop{\mathbb{E}}_{k_1',k_2'}\left(\mathrm{Tr}_M \sum_{i\neq(0^t,\Phi^+)} \langle i|\mu^{\mathsf{ideal}}_{k_1',k_2'}|i\rangle\right) = \mathrm{Tr}_M\left(\sum_{\ell\neq 0} \mathop{\mathbb{E}}_{\substack{k_1'\in\mathcal{K}_1\\ V_{k_1'}^\dagger P_\ell V_{k_1'}\notin\mathbb{P}_{\mathsf{ideal}}}} |\alpha_\ell|^2\, U_\ell^R \rho^{MR} U_\ell^\dagger\right) \tag{9}$$

$$\approx_\varepsilon \mathrm{Tr}_M\left(\sum_{\ell\neq 0} \mathop{\mathbb{E}}_{k_1'\in\mathcal{K}_1} |\alpha_\ell|^2\, U_\ell^R \rho^{MR} U_\ell^\dagger\right), \tag{10}$$

where the approximation sign means that the trace distance between the two states is upper bounded by ε. The closeness follows from the strong-purity-testing property of the code: the two states differ in those keys k_1' for which $V_{k_1'}^\dagger P_\ell V_{k_1'} \in \mathbb{P}_{\mathsf{ideal}} \subseteq \mathbb{P}_{\mathsf{real}}$, and for any non-identity Pauli P_ℓ, this set is small by strong purity testing. Combined with the facts that $\mathrm{tr}(U_\ell \rho U_\ell^\dagger) = 1$ and $\sum_\ell |\alpha_\ell|^2 = 1$, it follows that the states in equations (9) and (10) are ε-close. Note that none of the terms in equation (10) depends on k_1', so we can remove the expectation over it.

Applying the triangle inequality (twice), the second term in equation (7) is found to be small:

$$\mathop{\mathbb{E}}_{k_1}\frac{1}{2}\left\|\mathop{\mathbb{E}}_{k_1',k_2'}\left(\mathrm{Tr}_M \sum_{i\neq(0^t,\Phi^+)} \langle i|\mu^{\mathsf{ideal}}_{k_1',k_2'}|i\rangle\right) - \mathop{\mathbb{E}}_{k_2}\left(\mathrm{Tr}_M \sum_{i\neq 0^t} \langle i|\mu^{\mathsf{real}}_{k_1,k_2}|i\rangle\right)\right\|_{\mathrm{tr}} \tag{11}$$

$$\leq \frac{\varepsilon}{2} + \mathop{\mathbb{E}}_{k_1}\frac{1}{2}\left\|\mathrm{Tr}_M\left(\sum_{\ell\,:\,V_{k_1}^\dagger P_\ell V_{k_1}\in\mathbb{P}_{\mathsf{real}}\setminus\{\mathsf{I}^{\otimes(m+t)}\}} |\alpha_\ell|^2\, U_\ell \rho U_\ell^\dagger\right)\right\|_{\mathrm{tr}} \tag{12}$$

$$\leq \frac{\varepsilon}{2} + \frac{1}{2}\mathop{\mathbb{E}}_{k_1}\sum_{\ell\,:\,V_{k_1}^\dagger P_\ell V_{k_1}\in\mathbb{P}_{\mathsf{real}}\setminus\{\mathsf{I}^{\otimes(m+t)}\}} |\alpha_\ell|^2, \tag{13}$$

which we can upper bound by ε by applying the strong-purity-testing property once more.

Next, we bound the first term of equation (7): the difference between the ideal and the real channel in the accept case. The strategy is identical to the reject case that we just treated, but because we want to recycle both k_1 and k_2 in the accept case, we have to be more careful. The state in the real scenario, $\langle 0^t|\mu^{\mathsf{real}}_{k_1,k_2}|0^t\rangle$, cannot be rewritten into the compact form of, e.g., equation (8), because we cannot average over the Pauli key k_2. Using a technical lemma from [15] and Jensen's inequality in order to take the expectation over the keys inside, we obtain the bound

$$\mathop{\mathbb{E}}_{k_1,k_2}\left\|\mathop{\mathbb{E}}_{k_1',k_2'}\left(\langle\Phi^+,0^t|\mu^{\mathsf{ideal}}_{k_1',k_2'}|\Phi^+,0^t\rangle\right)-\langle 0^t|\mu^{\mathsf{real}}_{k_1,k_2}|0^t\rangle\right\|_{\mathrm{tr}}\leq\frac{\varepsilon}{2}+\sqrt{\varepsilon}. \tag{14}$$

For a complete derivation, see the full version.

We have now upper bounded $\frac{1}{2}\|\mathfrak{I}(\rho)-\mathfrak{R}(\rho)\|_{\mathrm{tr}}\ \leq\sqrt{\varepsilon}+\frac{3}{2}\varepsilon$ for any state ρ^{MRE}, resulting in $\frac{1}{2}\|\mathfrak{I}-\mathfrak{R}\|_\diamond\leq\sqrt{\varepsilon}+\frac{3}{2}\varepsilon$, as desired. ◀

4 A strong-purity-testing variation on the trap code

Theorem 9 already gives us a quantum-ciphertext-authenticating code with key recycling: the Clifford code. However, the Clifford code is not very well suited for quantum computing on authenticated data. Although all Clifford-group operations can be performed easily by updating the key client-side, it is not known how to perform non-Clifford gates and measurements on the Clifford code. Moreover, if an entangling operation is performed on two separately encoded qubits, their keys also have to be combined into a key for a single, bigger ciphertext. This prevents output qubits from being decrypted individually.

In this section, we therefore present a strong-purity-testing variation on the trap code, the *strong trap code*, which does allow for computation on the ciphertexts in a meaningful and efficient way. By Theorem 9, this construction immediately gives rise to a ciphertext authentication scheme with key recycling (QCA-R). Note that the strong trap code is also secure in Portmann's abstract-cryptography definition of quantum plaintext authentication with key recycling [15].

4.1 Benign distance and weight sparsity

The strong trap code requires the existence of a family of quantum error-correcting codes with two specific properties: a high benign distance, and weight sparsity. We specify these properties here.

If a QECC has distance d, it is not necessarily able to detect all Pauli errors of weight less than d. For example, if one of the qubits in a codeword is in the state $|0\rangle$, then a Pauli-Z remains undetected. In general, any Pauli error that stabilizes all codewords will remain undetected by the code. Of course, such an error does not directly cause harm or adds noise to the state, because it effectively performs the identity operation. However, in an adversarial setting, even such 'benign' Pauli errors indicate that someone tried to modify the state.

We consider an alternative distance measure for quantum error-correcting codes that describes the lowest possible weight of a stabilizer:

▶ **Definition 10** (Benign distance)**.** The *benign distance* of an $[[n,m]]$ code is the minimum weight of a non-identity Pauli P_ℓ such that $P_\ell|x_L\rangle=|x_L\rangle$ for all $x\in\{0,1\}^m$. If such P_ℓ does not exist, the benign distance is ∞.

To distinguish the benign distance from the notion of difference defined in Definition 4, we will often use the term *conventional distance* to refer to the latter.

The benign distance in a fixed relation to the conventional distance. For example, the $[[7,4]]$ Steane code has distance 3, but benign distance 4. On the other hand, the $[[49,1]]$ concatenated Steane code has distance 9, but a benign distance of only 4 (any non-identity stabilizer for the $[[7,4]]$ Steane code is also a stabilizer on the $[[49,1]]$ code if it is concatenated with identity on the other blocks). Even though the two quantities do not bound each other in general, we observe that the benign distance of weakly self-dual *CSS codes* (i.e., CSS codes constructed from a weakly self-dual classical code) grows with their conventional distance: see the full version.

We define a second property of interest: *weight sparsity*. Intuitively, weight sparsity means that for any set of X-, Y-, and Z-weights, randomly selecting a Pauli operator with those weights only yields a stabilizer with very small probability. This probability should shrink whenever the codeword length grows; for this reason, we consider weight sparsity as a property of code *families* rather than of individual codes.

▶ **Definition 11** (Weight-sparse code family)**.** Let $(E_i)_{i\in\mathbb{N}}$ be a family of quantum error-correcting codes with parameters $[[n(i), m(i), d(i)]]$. For each $i \in \mathbb{N}$, and for all non-negative integers x, y, z such that $x + y + z \leq n(i)$, let $A_i(x, y, z)$ denote the set of $n(i)$-qubit Paulis with X-weight x, Y-weight y, and Z-weight z. Let $B_i(x, y, z)$ denote set of benign Paulis in $A_i(x, y, z)$.

The family $(E_i)_{i\in\mathbb{N}}$ is *weight-sparse* if the function

$$f(i) := \max_{x+y+z\leq n(i)} \frac{|B_i(x,y,z)|}{|A_i(x,y,z)|}$$

is negligible[5] in $n(i)$.

In the full version of this paper, we construct a weight-sparse family of weakly self-dual CSS codes that have benign distance $O(\sqrt{n(i)})$, where $n(i)$ is the codeword length of the ith code in the family. The CSS codes are constructed from a punctured version of classical Reed–Muller codes [16].

4.2 The strong trap code

We present a modified version of the trap code, which we call the *strong trap code*. Contrary to the regular trap code, which appends $2t$ trap qubits, the strong trap code only appends a single $|0\rangle$ trap and a single $|+\rangle$ trap. These two traps are subsequently encoded using a quantum error-correcting code that has the desired properties described above, resulting in a ciphertext of the same length as the original trap code.

▶ **Definition 12** (Strong trap code)**.** Let $(E_i)_{i\in\mathbb{N}}$ be a weight-sparse family of weakly self-dual CSS codes with parameters $[[n(i), 1, d(i) = \Omega(\sqrt{n(i)})]]$ and benign distance $\Omega(\sqrt{n(i)})$. Then the ith strong trap code $\{V_{i,k}\}_{k\in\mathcal{K}_i}$ encodes $m = 1$ qubit using $t = 3n(i) - 1$ tags with the unitaries $V_{i,k} := \pi_k E_i^{\otimes 3} \mathsf{H}_{2n(i)+1}$ (where $\mathsf{H}_{2n(i)+1} = \mathsf{I}^{\otimes 2n(i)} \otimes \mathsf{H} \otimes \mathsf{I}^{\otimes(n(i)-1)}$).

[5] A function $f(x)$ is negligible in x if for all $c \in \mathbb{N}$, there exists an x_0 such that for all $x \geq x_0$, $f(x) < x^{-c}$. This definition is extended by stating that a function $f(x)$ is negligible in another function $g(x)$ if for all $c \in \mathbb{N}$, there exists an x_0 such that for all $x \geq x_0$, $f(x) < (g(x))^{-c}$.

The strong trap code invokes two layers of security: the CSS codes E_i, which detect low-weight attacks, and the traps $|0\rangle$ and $|+\rangle$, which detect higher-weight attacks by revealing bit and phase flips, respectively.

One can verify that computing on quantum states authenticated with the strong trap code works in much the same way as for the original trap code. For details, see [6].[6]

▶ **Theorem 13.** *The strong trap code is a strong-purity-testing code with error* $\mathrm{negl}(n(i))$.

Proof. Consider an arbitrary i and non-identity Pauli $P_\ell \in \mathbb{P}_{3n(i)} \backslash \{\mathsf{I}^{\otimes 3n(i)}\}$. Let w_x and w_z denote the X-weight and Z-weight (respectively) of P_ℓ, and note that $\max(w_x, w_z) > 0$.

We bound the probability that $P_{\ell'} := \pi_k^\dagger P_\ell \pi_k$ remains undetected by the code E_i and the traps. Because E_i is a CSS code, it detects X and Z errors separately: let us write $P_{\ell'} = P_x P_z$ with $P_x \in \{\mathsf{I}, \mathsf{X}\}^{\otimes 3n(i)}$ and $P_z \in \{\mathsf{I}, \mathsf{Z}\}^{\otimes 3n(i)}$, and focus first on the probability that P_x remains undetected, i.e., the probability that $\mathsf{H}_{2n(i)+1}(E_i^\dagger)^{\otimes 3} P_x E_i^{\otimes 3} \mathsf{H}_{2n(i)+1} \in \mathbb{P}_1 \otimes \{\mathsf{I}, \mathsf{Z}\}^{\otimes 3n(i)-1}$.

Because of the permutation π_k, P_x is a random Pauli in $\{\mathsf{I}, \mathsf{X}\}^{\otimes 3n(i)}$ with weight w_x. (Note that P_z is also a random Pauli with weight w_z, but is correlated with P_x: any overlap in the locations of X and Z operators in P_ℓ is preserved by the permutation.)

Consider all possible values of $w_x = w_1 + w_2 + w_3$, where w_1 denotes the weight of P_x on the first (data) codeword, w_2 the weight on the second ($|0\rangle$-trap) codeword, and w_3 the weight on the third ($|+\rangle$-trap) codeword:

- If $w_x = 0$, then the Pauli P_x is identity, and remains undetected with probability 1.
- If $0 < w_x < d(i)$, then $0 < w_j < d(i)$ for at least one $j \in \{1, 2, 3\}$. E_i detects an error on the jth block with certainty, since the weight of the error is below the distance and the benign distance.
- If $d(i) \leq w_x \leq 3n(i) - d(i)$, the attack P_x will likely be detected on the second block, the $|0\rangle$-trap. We can be in one of four cases:
  - $w_2 > 0$ and P_x is detected in the second block by the CSS code E_i.
  - $w_2 > 0$ and P_x acts as a logical operation on the second block. Since P_x consists of only I's and X's, this logical operation can only be an X by the construction of CSS codes. In this case, P_x is detected by the projection that checks whether the trap is still in the $|0\rangle$ state.
  - $w_2 > 0$ and P_x acts as a stabilizer on the second block, and remains undetected on that block. However, by the weight-sparsity of the code family, the probability that this is the case is negligible in $n(i)$.
  - $w_2 = 0$. In this case, P_x acts as identity on the second block. The probability that this case occurs, however, is small:

$$\Pr_k[w_2 = 0] = \frac{\binom{2n(i)}{w_x}}{\binom{3n(i)}{w_x}} < \left(\frac{2}{3}\right)^{w_x} \leq \left(\frac{2}{3}\right)^{d(i)} . \tag{15}$$

    The first inequality holds in general for binomials, and the second one follows from the fact that $w_x \geq d(i)$. Since $d(i) = \Omega(\sqrt{n(i)})$, this probability is negligible in $n(i)$.

  In total, the probability of the attack remaining undetected for $d(i) \leq w_x \leq 3n(i) - d(i)$ is negligible in $n(i)$.

[6] For some applications, authenticating through measurement (cf. [6, Appendix B.2]) can be very useful. Our underlying code has all the requirements to achieve this in principle, but in this work we focus on quantum authentication and do not formulate the full security notions needed to properly describe this scenario.

- If $3n(i) - d(i) < w_x < 3n(i)$: as in the second case, there is at least one $j \in \{1,2,3\}$ such that $n(i) - d(i) < w_j < n(i)$, causing the attack to be detected (recall that $\mathsf{X}^{\otimes 3n(i)}$ is a logical X, and therefore this mirrors the $0 < w_x < d(i)$ case).
- If $w_x = 3n(i)$, then the logical content of the second block, the $|0\rangle$-trap, is flipped. This is detected with certainty as well.

We see that unless $w_x = 0$, the Pauli P_x remains undetected only with probability negligible in $n(i)$. A similar analysis can be made for P_z: it is always detected with high probability, unless $w_z = 0$. We stress that these probabilities are *not* independent. However, we can say that

$$\Pr_k[P_x \text{ and } P_z \text{ undetected}] \leq \min\left\{\Pr_k[P_x \text{ undetected}],\ \Pr_k[P_z \text{ undetected}]\right\}, \tag{16}$$

and since at least one of w_x and w_z is non-zero, this probability is negligible in $n(i)$. ◀

5 Simultaneous encryptions with key reuse

Earlier work on key reuse for quantum authentication deals explicitly with *key recycling*, the decision to reuse (part of) a key for a new encryption after completing the transmission of some other quantum message. The key is reused only *after* the honest party decides whether to accept or reject the first message, so recycling is a strictly sequential setting.

If Construction 1 is instantiated with a strong-purity-testing code (such as the strong trap code), the resulting scheme is able to handle an even stronger, parallel, notion of key reuse. As long as the one-time pads are independent, it is possible to encrypt multiple qubits under the same code key while preserving security. Even if the adversary is allowed to interactively decrypt a portion of the qubits one-by-one, the other qubits will remain authenticated. This property is especially important for the strong trap code: computing on data authenticated with the strong trap code requires all qubits to be encrypted under the same permutation key.

The original trap code is secure in this setting (as long as the one-time pads are fresh; see Section 5.2 of [6]), but only if all qubits are decrypted at the same time. If some qubits can be decrypted separately, the adversary can deduce the location of the $|+\rangle$ traps by applying single-qubit X operations to different ciphertexts at different locations, and observing which ones are rejected. Repeating this for the Z operator to learn about the $|0\rangle$ traps, the adversary can completely break the authentication on the remaining qubits.

Suppose we encrypt two messages using an authentication scheme based on a strong-purity-testing code $\{V_{k_0}\}_{\mathcal{K}_0}$, using the same code key k_0 but a fresh one-time pad. If we then decrypt the first message, the scheme is still QCA-R-authenticating on the second message with only slightly worse security.

▶ **Theorem 14** (informal). *Let* (Encrypt, Decrypt) *be an* ε-QCA-R-*authenticating scheme resulting from Construction 1, using a strong-purity-testing code* $\{V_{k_0}\}_{\mathcal{K}_0}$. *Let* M_1, M_2 *denote the plaintext registers of the two messages,* $C_1 = M_1T_1, C_2 = M_2T_2$ *the corresponding ciphertext registers, and* R *a side-information register. Let* $\mathcal{A}_1$, $\mathcal{A}_2$ *be arbitrary adversarial channels. Consider the setting where the adversary acts on the qubits, encrypted with keys* k_0, k_1, k_2, *as*

$$\mathsf{Decrypt}_{k_0,k_2}^{C_2\to M_2} \circ \mathcal{A}_2^{M_1,C_2,R} \circ \mathsf{Decrypt}_{k_0,k_1}^{C_1\to M_1} \circ \mathcal{A}_1^{C_1,C_2,R} \circ \left(\mathsf{Encrypt}_{k_0,k_1}^{M_1\to C_1} \otimes \mathsf{Encrypt}_{k_0,k_2}^{M_2\to C_2}\right),$$

so that the key k_0 *is used for both messages. Then, the scheme is* 2ε-QCA-R-*authenticating on the second qubit.*

Sketch. As a first step, we rewrite the encryption of the second qubit as using encoding and teleportation, by using the equivalence between applying a random quantum one-time pad and teleporting a state. The encryption of the second qubit can then be thought of as happening after decryption of the first qubit. Next, we apply QCA-R security of the first qubit, where we are using the property that k_0 is recycled both in the accept and the reject case. Finally we undo the rewrite and can directly apply QCA-R security on the remaining state. ◀

The complete proof can be found in the full version. The argument easily extends to any polynomial number of authenticated qubits.

6 Conclusion

We presented a new security definition, QCA-R, for ciphertext authentication with key recycling, and showed that schemes based on purity-testing codes satisfy quantum ciphertext authentication, while strong purity testing implies both ciphertext authentication and key recycling. This is analogous to the security of quantum plaintext-authentication schemes from purity-testing codes [5, 15].

Additionally, we constructed the *strong trap code*, a variant of the trap code which is a strong-purity-testing code and therefore is QCA-R secure (as well as secure under all notions of plaintext authentication). This new scheme can strengthen security and add key-recycling to earlier applications of the trap code. It is also applicable in a wider range of applications than the original trap code, because encrypted qubits remain secure even if other qubits sharing the same key are decrypted earlier.

A potential application of the strong trap code is the design of a quantum CCA2-secure encryption scheme (as in [3, Definition 9]) that allows for computation on the encrypted data. By only using the pseudo-random generator for the one-time-pad keys, and recycling the key for the underlying error-correcting code, this security level could be achieved.

As future work, our definition of QCA-R could be generalized in different ways. First, one can consider a variant of the definition in the abstract-cryptography or universal-composability framework, in order to ease the composition with other cryptographic primitives. Second, because it can be useful to authenticate measurements in delegated computation applications, one could extend the definition of QCA-R to deal with the measurement of authenticated data. We expect no real obstacles for this extension of the definition, and refer to [6, Appendix B.2] for comparable work on the original trap code.

References

1 Dorit Aharonov, Michael Ben-Or, and Elad Eban. Interactive proofs for quantum computations. *arXiv preprint arXiv:0810.5375*, 2008.

2 Gorjan Alagic, Yfke Dulek, Christian Schaffner, and Florian Speelman. Quantum fully homomorphic encryption with verification. In *Advances in Cryptology – ASIACRYPT 2017*, pages 438–467, Cham, 2017. Springer International Publishing. doi:10.1007/978-3-319-70694-8_16.

3 Gorjan Alagic, Tommaso Gagliardoni, and Christian Majenz. Unforgeable quantum encryption. *arXiv preprint arXiv:1709.06539*, 2017.

4 Gorjan Alagic and Christian Majenz. Quantum non-malleability and authentication. In *Advances in Cryptology – CRYPTO 2017*, pages 310–341, Cham, 2017. Springer International Publishing. doi:10.1007/978-3-319-63715-0_11.

5 Howard Barnum, Claude Crépeau, Daniel Gottesman, Adam Smith, and Alain Tapp. Authentication of quantum messages. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 449–458. IEEE, 2002.

6 Anne Broadbent, Gus Gutoski, and Douglas Stebila. Quantum one-time programs. In *Advances in Cryptology–CRYPTO 2013*, pages 344–360. Springer, 2013.

7 Anne Broadbent, Zhengfeng Ji, Fang Song, and John Watrous. Zero-knowledge proof systems for QMA. In *57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 31–40, Oct 2016. doi:10.1109/FOCS.2016.13.

8 Anne Broadbent and Evelyn Wainewright. Efficient simulation for quantum message authentication. In *Information Theoretic Security*, pages 72–91, Cham, 2016. Springer International Publishing. doi:10.1007/978-3-319-49175-2_4.

9 Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009.

10 Frédéric Dupuis, Jesper Buus Nielsen, and Louis Salvail. Actively secure two-party evaluation of any quantum operation. In *Advances in Cryptology – CRYPTO 2012*, volume 7417, pages 794–811. Springer International Publishing, 2012. Full version on IACR eprint archive: eprint.iacr.org/2012/304. doi:10.1007/978-3-642-32009-5_46.

11 Serge Fehr and Louis Salvail. Quantum authentication and encryption with key recycling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 311–338. Springer, 2017.

12 Sumegha Garg, Henry Yuen, and Mark Zhandry. New security notions and feasibility results for authentication of quantum data. In Jonathan Katz and Hovav Shacham, editors, *Advances in Cryptology – CRYPTO 2017*, pages 342–371, Cham, 2017. Springer International Publishing. doi:10.1007/978-3-319-63715-0_12.

13 Patrick Hayden, Debbie W Leung, and Dominic Mayers. The universal composable security of quantum message authentication with key recyling. *arXiv preprint arXiv:1610.09434*, 2016.

14 Jonathan Oppenheim and Michał Horodecki. How to reuse a one-time pad and other notes on authentication, encryption, and protection of quantum information. *Phys. Rev. A*, 72:042309, Oct 2005. doi:10.1103/PhysRevA.72.042309.

15 Christopher Portmann. Quantum authentication with key recycling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 339–368. Springer, 2017.

16 John Preskill. Quantum computation, 1997. URL: http://www.theory.caltech.edu/people/preskill/ph229/index.html.

17 Mark N. Wegman and J. Lawrence Carter. New hash functions and their use in authentication and set equality. *Journal of Computer and System Sciences*, 22(3):265–279, 1981. doi:10.1016/0022-0000(81)90033-7.